

Sistemi Operativi

11 giugno 2012

Compitino 2

Si risponda ai seguenti quesiti, giustificando le risposte.

1. Spiegare con esempi perché la dimensione del working set di un processo si può incrementare o decrementare durante la sua esecuzione.

Risposta: (5 punti) La dimensione del working set può cambiare durante l'esecuzione di un processo in quanto si può verificare una successione di fasi di esecuzione che coinvolgono un numero differente di pagine. Ad esempio, si consideri un programma con due cicli:

1. il primo ciclo esegue delle operazioni sugli elementi di un piccolo vettore;
2. il secondo ciclo opera una sommatoria degli elementi di un vettore di dimensione maggiore.

A questo punto si può concludere che il working set del processo sarà limitato (perché avrà bisogno di accedere a poche pagine) fintanto che eseguirà le istruzioni all'interno del primo ciclo, mentre crescerà (dato che il secondo vettore è di dimensione maggiore e quindi occuperà più pagine) quando il processo passerà ad eseguire le istruzioni del secondo ciclo.

Altri esempi di processi con working set di dimensione variabile sono i compilatori di linguaggi che suddividono in fasi la propria esecuzione (analisi lessicale, parsing, costruzione della tabella dei simboli, ecc.).

2. In un sistema che usa paginazione, l'accesso al TLB richiede 150ns, mentre l'accesso alla memoria richiede 400ns. Quando si verifica un page fault, si perdono 8ms per caricare la pagina che si sta cercando in memoria. Se il page fault rate è il 2% e il TLB hit il 70%, indicare l'EAT ai dati.

Risposta: (6 punti) Convertendo tutti i tempi in ms, abbiamo:

- tempo di accesso al TLB: $150 \text{ ns} = 150 \cdot 10^{-6} \text{ ms}$;
- tempo di accesso alla memoria: $400 \text{ ns} = 400 \cdot 10^{-6} \text{ ms}$;
- tempo di gestione del page fault: 8 ms ;
- page fault rate: $2\% = 0,02$;
- TLB hit: $70\% = 0,7$.

Quindi:

$$\begin{aligned} EAT &= \text{TLB hit} \cdot (150 \cdot 10^{-6} + 400 \cdot 10^{-6}) + (1 - \text{TLB hit}) \cdot [(1 - \text{page fault rate})(150 \cdot 10^{-6} + 2 \cdot 400 \cdot 10^{-6}) + \text{page fault rate} \cdot (150 \cdot 10^{-6} + 8 + 2 \cdot 400 \cdot 10^{-6})] \\ &= 0,7 \cdot (150 \cdot 10^{-6} + 400 \cdot 10^{-6}) + 0,3 \cdot [0,98 \cdot (150 \cdot 10^{-6} + 800 \cdot 10^{-6}) + 0,02 \cdot (150 \cdot 10^{-6} + 8 + 800 \cdot 10^{-6})] \\ &= 0,7 \cdot (5,5 \cdot 10^{-4}) + 0,3 \cdot [0,98 \cdot 9,5 \cdot 10^{-4} + 0,02 \cdot 8,00095] \\ &= 3,85 \cdot 10^{-4} + 0,3 \cdot (9,31 \cdot 10^{-4} + 0,160019) = 3,85 \cdot 10^{-4} + 2,793 \cdot 10^{-4} + 0,0480057 \\ &= 0,04867 \text{ ms} \end{aligned}$$

3. Si spieghi come funziona l'allocazione contigua, specificando lo schema di traduzione dell'indirizzo logico in indirizzo fisico. Questa soluzione soffre di frammentazione esterna? (in caso di risposta affermativa spiegare perché).

Risposta: (5 punti) Lo schema di allocazione contigua prevede che ogni file occupi un insieme di blocchi contigui sul disco. È quindi molto semplice da implementare in quanto basta conoscere l'indirizzo del blocco iniziale e la lunghezza per accedere ai blocchi del file. La traduzione da indirizzo logico a fisico (indicando con B la dimensione del blocco e LA l'indirizzo logico) avviene eseguendo la divisione intera di LA per B : il quoziente va sommato all'indirizzo del blocco iniziale per determinare l'indirizzo del blocco da accedere, mentre il resto indicherà l'offset all'interno di quest'ultimo.

L'allocazione contigua soffre di frammentazione esterna in quanto, dovendo i file occupare una sequenza di blocchi contigui, in seguito a creazioni e cancellazioni di file si può avere una situazione in cui lo spazio libero totale su disco è sufficiente a soddisfare una nuova richiesta, ma non è contiguo, ovvero, i "buchi" di blocchi liberi contigui sono troppo piccoli per la richiesta in oggetto.

4. (a) Si spieghi cosa si intende per sistema di dischi RAID.
(b) Si illustri come è organizzato il livello RAID 1, evidenziandone pregi e svantaggi.

Risposta:

Sistemi Operativi

11 giugno 2012

Compitino 2

- (a) (2 punti) L'architettura RAID è stata introdotta per migliorare l'affidabilità e le prestazioni del sistema dei dischi in un sistema di calcolo. Infatti, aumentando il numero di dischi fisici presenti in un calcolatore, diminuisce proporzionalmente il tempo medio fra un guasto e l'altro. Quindi il sistema RAID introduce degli schemi di ridondanza dei dati per consentire il funzionamento dei dischi anche in situazioni critiche.
- (b) (2 punti) Al livello 1 (mirroring o shadowing) i contenuti di ogni disco sono interamente duplicati in un altro disco del sistema, garantendo un'eccellente resistenza ai crash, alte performance in lettura, ma non in scrittura (dove le prestazioni sono le stesse di un sistema non-RAID).

5. Si descrivano i passi principali eseguiti dal driver delle interruzioni.

Risposta: (5 punti) I passi eseguiti dal driver delle interruzioni sono i seguenti:

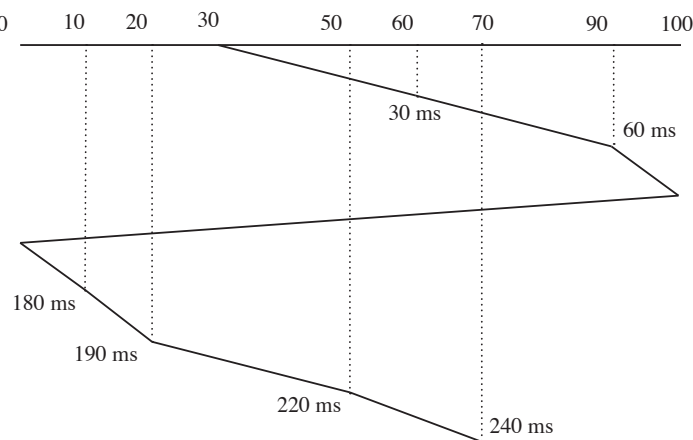
- salvare i registri della CPU,
- impostare un contesto per la procedura di servizio (inizializzare TLB, MMU, stack ecc.),
- inviare un segnale di *acknowledge* al controllore degli interrupt (per avere interrupt annidati),
- copiare la copia dei registri nel PCB,
- eseguire la procedura di servizio che accede al dispositivo,
- eventualmente, cambiare lo stato a un processo in attesa (e chiamare lo scheduler di breve termine),
- organizzare un contesto (TLB, MMU ecc.) per il processo successivo,
- caricare i registri del nuovo processo dal suo PCB,
- continuare l'esecuzione del processo selezionato.

6. Si consideri un disco con un intervallo di tracce da 0 a 100, gestito con politica C-SCAN (SCAN circolare) con direzione di servizio verso tracce con numero crescente. Inizialmente la testina è posizionata sul cilindro 30; lo spostamento ad una traccia adiacente richiede 1 ms. Al driver di tale disco arrivano richieste per i cilindri 90, 20, 60, 70, 50, 10, rispettivamente agli istanti 0 ms, 10 ms, 20 ms, 80 ms, 90 ms, 150 ms. Si trascuri il tempo di latenza.

1. In quale ordine vengono servite le richieste?
2. Il tempo di attesa di una richiesta è il tempo che intercorre dal momento in cui è sottoposta al driver a quando viene effettivamente servita. Qual è il tempo di attesa medio per le sei richieste in oggetto?

Risposta:

1. (3 punti) Le richieste vengono servite nell'ordine 60, 90, 10, 20, 50, 70:



2. (2 punti) Il tempo di attesa medio per le sei richieste in oggetto è

$$\frac{(30-20)+(60-0)+(180-150)+(190-10)+(220-90)+(240-80)}{6} = \frac{10+60+30+180+130+160}{6} = \frac{570}{6} = 95 \text{ ms.}$$